

Desarrollo de ecuaciones predictivas para el cálculo de composición corporal por impedanciometría

D. Bellido, J. Carreira

Endocrinología y Nutrición, Hospital Arquitecto Marcide, Ferrol,

Correspondencia: Dr. Diego Bellido Guerrero. C/ Hospital nº 29-31 (4º -A), 15401-Ferrol, La Coruña

E-mail: diego_bellido@arrakis.es

Resumen

El análisis de la composición corporal se ha convertido en un elemento básico dentro de la exploración clínica tanto inicial, como en el seguimiento evolutivo de aquellas patologías en las que las variaciones de los distintos compartimentos y componentes del organismo representan la aparición de factores de riesgo adicionales o influyen en la respuesta a los diferentes tratamientos.

En el caso particular de la obesidad, como patología de carácter crónico, la determinación de los compartimentos graso, con sus diferentes componentes, y no graso tiene un papel central en la evaluación inicial y el seguimiento del paciente obeso.

Las técnicas que miden de forma muy aproximada los diferentes componentes del organismo no están disponibles para su uso en la práctica clínica diaria, y deben emplearse métodos que miden algún componente o propiedad corporal y, mediante el desarrollo de fórmulas matemáticas estimar los diferentes compartimentos corporales.

El desarrollo de ecuaciones matemáticas basadas en el análisis por impedancia bioeléctrica permite disponer de fórmulas específicas para diferentes poblaciones según edad, sexo, índice de masa corporal, etc., aplicables en el contexto clínico diario.

Para la construcción de estas fórmulas matemáticas se emplea la regresión lineal múltiple, y de la adecuada elección de las variables, del correcto ajuste del modelo y su posterior validación dependerá la obtención de ecuaciones que puedan ser de utilidad en los diferentes campos clínicos y epidemiológicos.

En este artículo nos proponemos revisar la metodología de desarrollo de ecuaciones de regresión lineal múltiple aplicada al análisis de composición corporal por bioimpedancia eléctrica y su posterior validación con técnicas de referencia mediante diferentes métodos gráficos y numéricos.

Summary

Corporal composition analysis is an important tool in the initial approach and follow-up of pathologies in which variations in different compartments and compounds of body represent additional risk factors or influence the response to different treatments.

In the particular case of obesity, a chronic disease, the determination of fat mass and fat free mass is fundamental in the initial and follow up evaluation of obese patients.

Techniques that measure the different compartments are not available in the real clinical practice; we must use methods that measure one compartment or characteristic of the body, and by means of the development of mathematic formulas to estimate the different body compartments.

The development of mathematical equations based in bio-impedance analysis adapted to different populations adjusted by age, sex, body mass index, etc., is useful in the clinical practice.

Multiple regression analysis is used to develop these equations; the right election of variables, adjusted model, and post hoc validation may produce equations useful in clinic and epidemiological areas.

In this article, we review the methodology for the development of multiple regression equations for the analysis of corporal composition with bio-impedance, and the validation of these equations with gold Standard methods using graphic and numeric approaches.

Introducción

El estudio de la composición corporal en humanos dirigida a determinar componentes graso, no graso, líquidos extra e intracelulares y masa ósea, tiene como objetivo detectar aumentos o deficiencias que puedan generar patologías que impliquen aumento en la morbimortalidad, así como seguir la evolución de determinadas patologías y su respuesta al tratamiento.

Los intentos de establecer las proporciones ideales del cuerpo humano se remontan al siglo XIX, lo que dio origen a la antropometría; diversos estudios posteriores desarrollaron y emplearon instrumentos como los lipocalibres destinados a cuantificar los pliegues subcutáneos en un intento de estimar el contenido graso corporal. Las diversas mediciones efectuadas por instrumentos usados en antropometría como báscula, cinta métrica para circunferencias corporales, estadiómetro y lipocalibre originaron las primeras ecuaciones matemáticas en un intento de integrar todas estas medidas en fórmulas que permitieran aproximarse al contenido graso y no graso del cuerpo; estas técnicas siguen utilizándose en la actualidad dentro de la valoración inicial del estado nutricional; como ejemplo podemos citar el conocido índice de masa corporal (IMC), resultado de dividir el peso en kg por la estatura (expresada en m) al cuadrado, que se ha comprobado que guarda una alta correlación con el contenido de grasa corporal.

Desde hace varias décadas se han venido desarrollando métodos que permitieran analizar la composición corporal de forma fiable y reproducible y que, al mismo tiempo sean sencillos y rápidos de aplicar en la clínica diaria y en la investigación epidemiológica, ya que los métodos que miden la com-

posición corporal con mayor exactitud, los considerados de referencia, como el caso de la hidrodensitometría y el análisis de activación de neutrones, están reservados para la investigación debido a factores como coste, tiempo requerido para el análisis, sofisticación en el manejo, infraestructura requerida, etc.

Los métodos de estudio de la composición corporal se clasifican en dos tipos, según su objetivo sea medir componentes o propiedades del cuerpo humano. Los métodos basados en propiedades miden una característica del cuerpo humano; ejemplos son la hidrodensitometría que mide la densidad corporal y la impedancia bioeléctrica que cuantifica la resistencia corporal al paso de una corriente eléctrica. Asumiendo unas relaciones preestablecidas entre propiedades y componentes que constituyen el organismo, permiten obtener fórmulas matemáticas que cuantifican el componente a partir de la propiedad medida.¹

Impedancia bioeléctrica y composición corporal

La impedancia bioeléctrica como método de estudio de la composición corporal ha tenido un desarrollo cada vez mayor en los últimos años, debido a que reúne una serie de características que hacen que sea de especial utilidad en clínica diaria y en estudios poblacionales de composición corporal, como son la portabilidad, bajo coste, mínimas molestias y riesgos para el paciente, mínimo entrenamiento para el personal operador, lo que contribuye a reducir los errores inter e intraobservador, etc. Además, se ha demostrado buena correlación con técnicas de referencia.

La impedancia bioeléctrica mide la resistencia que oponen las células y los fluidos corporales al paso de una corriente eléctrica alterna, a través de unos electrodos dispuestos en la superficie cutánea, que varían su colocación dependiendo del tipo de análisis que queramos realizar. Este valor de impedancia bioeléctrica, integrado con el sexo y valores antropométricos seleccionados, como talla, peso, circunferencias corporales y pliegues cutáneos, se utiliza para la predicción de la composición corporal a partir de ecuaciones desarrolladas mediante técnicas estadísticas multivariantes de regresión lineal, que nos permitirán obtener el valor de la masa libre de grasa o el porcentaje de masa grasa, así como el valor del agua corporal total, subdividiendo ésta en sus componentes extra e intracelular, estimados por métodos específicos de impedancia bioeléctrica.

La ecuación de regresión así obtenida pasa a formar parte del *software* del aparato de impedancia bioeléctrica, para proporcionarnos los valores de masa grasa, masa libre de grasa y agua corporal total a partir de los datos antropométricos que nosotros introduzcamos al realizar el análisis de impedancia bioeléctrica. Alternativamente, estos valores antropométricos y el valor de la impedancia bioeléctrica nos permitirán desarrollar ecuaciones de regresión específicas para poblaciones determinadas.

En este artículo nos proponemos revisar el método de desarrollo de ecuaciones de regresión para su aplicación con el análisis de impedancia bioeléctrica en el estudio de la composición corporal.

Consideraciones metodológicas

En general, las ecuaciones de regresión presentan dos tipos de variables, la variable dependiente, que es la que pretendemos medir, y las variables independientes o predictoras, relacionadas con la dependiente, que servirán para desarrollar las ecuaciones que permitan estimar el valor de la variable dependiente.

Como hemos mencionado previamente, se han estudiado diversas variables antropométricas que, junto al valor de la impedancia bioeléctrica, permiten obtener un análisis de la composición corporal basado en un modelo bicompartimental de

masa libre de grasa y masa grasa; las variables como sexo, peso, talla, y las variables que proporcionan medidas de la distribución regional de la grasa, como son los pliegues cutáneos (tricipital, bicipital) y las circunferencias corporales (cintura, cadera, muslo) tienen importancia como variables predictoras dependiendo de su relación biológica y de su correlación estadística con la variable que pretendemos medir, en nuestro caso la masa libre de grasa y el porcentaje de masa grasa.

En el momento de desarrollar una ecuación que permita estimar un parámetro determinado de la composición corporal se deben seleccionar las variables predictoras siguiendo criterios biológicos y no solo estadísticos; de tal forma que aunque una variable no presente una correlación estadísticamente significativa con la variable dependiente debería ser incluida en el modelo si consideramos que puede influir en la validez de la ecuación final,² como por ejemplo el caso de la variable sexo.

En cuanto a las variables predictoras a incluir en el modelo, al desarrollar la ecuación debemos seguir ciertos criterios según la variable que pretendemos estimar;^{2,3} así los pliegues corporales, al ser una medida del contenido adiposo subcutáneo deberían ser incluidas en las ecuaciones cuando la variable dependiente es el porcentaje de masa grasa; por otro lado, si lo que pretendemos estimar es la masa libre de grasa, debemos incluir variables antropométricas que en los diferentes estudios han demostrado buena correlación con la masa libre de grasa como son la estatura, que junto con el peso también es buen predictor del contenido corporal de masa grasa, y las medidas de la resistencia y reactancia corporales; en el caso de los análisis de impedancia bioeléctrica monofrecuencia realizados a 50 kHz, pueden superponerse los valores de la impedancia y resistencia, al ser la reactancia mínima.

La estatura y la resistencia se han revelado, a través de numerosos estudios, como las variables que presentan una mejor correlación estadística con la masa libre de grasa,^{3,4} por lo que forman parte, como variables predictoras, de la mayoría de las ecuaciones publicadas en la literatura, bien como términos independientes, o bien formando parte del denominado índice de resistencia: T^2/R (siendo T la medida de la estatura en cm y R la resistencia en ohmios). En la Tabla 1 se relacionan algunas de las variables predictoras empleadas con más frecuen-

Tabla 1. Variables predictoras empleadas en la literatura para estimar la masa libre de grasa y la masa grasa

Variable dependiente	Variables predictoras	Autor
Masa libre de grasa	T ² /R, peso, Xc, sexo	Kyle <i>et al.</i> ⁵
Masa libre de grasa	T ² /R, peso, Xc	Sun <i>et al.</i> ⁶
Masa libre de grasa	T ² /R, peso, edad	Segal <i>et al.</i> ⁷
Masa libre de grasa	T ² /R, T, peso, edad, sexo	Deurenberg <i>et al.</i> ⁸
Masa libre de grasa	R, T ² , peso, Xc, edad	Stolarczyk <i>et al.</i> ⁹
Masa libre de grasa	T ^{1.48} /R ^{0.55} , peso	Kotler <i>et al.</i> ¹⁰
Masa libre de grasa	T ² /R, peso, Xc	Roubenoff <i>et al.</i> ¹¹
Masa de grasa	T ² /R, peso, T, peso-sexo, edad	Heitmann ¹²

T: talla en cm, R: resistencia a 50 KHz, Xc: reactancia.

cia en las ecuaciones publicadas en la literatura para estimar la masa libre de grasa y la masa grasa.

Métodos estadísticos

Los objetivos que se persiguen con el desarrollo de ecuaciones de regresión son de dos tipos:

1. Obtener modelos que pretendan predecir el valor de una variable en función del valor que tomen una o varias variables; son los llamados modelos predictivos. A este modelo pertenecen las ecuaciones de composición corporal que utilizan la impedancia bioeléctrica.
2. Determinar y cuantificar la relación entre determinadas variables y la variable dependiente, también llamada variable respuesta; en este caso se trata de los modelos explicativos; ejemplos de este modelo son las ecuaciones que se utilizan para investigar qué variables afectan a parámetros fisiológicos, o modelos que intentan predecir el desarrollo de ciertas patologías en función de ciertos factores considerados de riesgo.

Ambos modelos difieren en los factores que deben tenerse en cuenta durante el desarrollo de las ecuaciones y que pueden afectar a la fiabilidad final del modelo, como son las relaciones entre variables independientes y dependiente, y de las independientes entre sí.

La aparición de la informática, y el desarrollo de paquetes informáticos cada vez más completos, ha facilitado enormemente el manejo de volúmenes elevados de datos y los cálculos matemáticos que requieren técnicas complejas como son las técnicas estadísticas multivariantes empleadas en el desarrollo de las ecuaciones de regresión, con el consiguiente avance en las técnicas de composición corporal que emplean esta metodología.

Propiedades de las variables

El método estadístico empleado para el desarrollo de las ecuaciones de composición corporal por impedancia bioeléctrica es el análisis de regresión lineal por el método de los mínimos cuadrados que, previo a cualquier tipo de análisis, requiere el cumplimiento de una serie de condiciones, por parte de las variables seleccionadas para la construcción del modelo.^{3,4} Estas condiciones son:

- a) **Linealidad:** La relación de las variables predictoras con la variable que pretendemos estimar debe ser lineal. Para confirmar esto previamente a cualquier tipo de análisis estadístico se debería realizar un gráfico de correlación que mostrara los valores de la variable dependiente con los de cada una de las variables predictoras. Las nubes de puntos resultantes deberían asemejarse a una línea recta, con mayor o menor dispersión a su alrededor. Las variables predictoras que no guarden relación lineal con la variable dependiente deberían ser transformadas para evitar errores de estimación cuando la ecuación se aplique a muestras independientes. Posteriormente, en el análisis ya cuantificaremos esta correlación mediante el coeficiente de correlación lineal de Pearson.
- b) **Homocedasticidad:** Este término hace referencia a que los residuales (diferencia entre valores de la variable dependiente medidos por la técnica de referencia y los valores estimados por la ecuación), deben seguir una distribución normal de media 0 y varianza 1. Para analizar si se cumple esta condición se pueden emplear test como los contrastes de Levene, Cochran o Barlett; o bien realizar un gráfico de correlación en el que se enfrentan los valores de la variable dependiente estimados por el modelo contra los residuales.^{2,13}

La nube de puntos resultante debe ser paralela al eje de abscisas (si el valor de la variable lo hemos colocado en este eje) y seguir una distribución aleatoria a ambos lados de una línea recta, sin mostrar ninguna tendencia. En el caso de que esto no se cumpla, por ejemplo, en los casos que para valores altos de la variable dependiente se produzcan residuales siempre del mismo signo y cada vez más altos, se dice que existe heterocedasticidad. Las causas de falta de homocedasticidad pueden ser la falta de linealidad o la presencia de valores extremos, que de confirmarse deben ser tratados según se comenta en sus respectivos apartados.

- c) *Normalidad*: Como mencionamos en el párrafo anterior, los residuales deben seguir una distribución normal, que se comprueba mediante los test de Kolmogorov-Smirnov, con la corrección de Lilliefors, o bien, gráficamente, podemos aproximarnos mediante los histogramas, gráficos de cajas y otras opciones incluidas en los programas estadísticos. Si esta condición no se cumple, puede deberse a asimetría en la distribución de la variable dependiente por la excesiva prevalencia de un grupo de valores. En ocasiones, esto también ocasiona problemas de falta de linealidad y heterocedasticidad; en este caso debemos intentar una transformación de la variable dependiente que solucionará ambos problemas; las más empleadas son las transformaciones Box-Cox.³
- d) *Ausencia de colinealidad*: Cuando el número de variables predictoras incluidas en el modelo es elevado se corre el riesgo de que algunas tengan entre sí un alto coeficiente de correlación lineal, lo que equivaldría a que alguna de estas variables pudiera ser, a su vez, estimada por otras también incluidas; esto reduce la precisión de los coeficientes de regresión estimados resultando en una falta de fiabilidad del modelo cuando se aplique a muestras diferentes a la usada para desarrollarlo, aunque obtengamos un elevado coeficiente de regresión múltiple R^2 , también llamado coeficiente de determinación. La presencia de colinealidad puede detectarse al realizar la regresión en el valor que toma el llamado factor de incremento de la varianza.¹³ La colinealidad puede afectar a unas variables predictoras y a otras no, y se produce cuando intentamos obtener de la muestra más información de la que

realmente puede proporcionarnos. Este problema puede evitarse eliminando del modelo variables que aporten información ya incluida en otras variables seleccionadas; a este concepto se refiere el término de parsimonia en regresión lineal múltiple, entendido como el intento de obtener siempre el mejor de los modelos con el menor número posible de variables predictoras. Volveremos a hablar de este concepto al hablar de la elección del modelo.

Para detectar las variables intercorrelacionadas posemos usar los gráficos matriciales de dispersión y las matrices de correlación de variables predictoras que nos proporcionan los programas de análisis estadístico.

- e) *Valores extremos o atípicos*: Al desarrollar una ecuación de regresión sobre una muestra de datos debemos prestar atención a los valores extremos que puedan existir en las variables predictoras. Estos casos, detectados al observar los gráficos de correlación de cada una de las variables predictoras frente a la variable dependiente, deben analizarse para determinar su influencia en el modelo y decidir si deben extraerse y ser sometidos a un análisis diferenciado, o si permanecen en la muestra que se usará para desarrollarlo. En este caso deberían emplearse los llamados métodos robustos de regresión que, a diferencia del método general, resisten la influencia de los valores atípicos.¹³

Si la causa de los valores extremos se debe a errores en la medición de las variables predictoras, evidentemente se corregirá sustituyendo por el valor correcto, y si no fuese posible se eliminarían de la muestra. También debe tenerse en cuenta que valores extremos en dos variables predictoras pueden ser causa de colinealidad entre ellas, por lo que al suprimir los valores atípicos desaparecerá la colinealidad.

Factores como la interacción y la detección de variables de confusión tienen importancia en el desarrollo de modelos explicativos; en el tipo de modelo de ecuación que nos ocupa, de tipo predictivo, interesa hallar una ecuación que estime lo más realmente posible el valor de la variable dependiente, por ejemplo la masa libre de grasa, siendo totalmente secundaria la posibilidad de interacción entre variables predictoras.

Validez del modelo

La validez de una ecuación de regresión en la predicción de una variable de composición corporal depende de varios factores, como son: validez de la variable dependiente, precisión en las medidas de las variables independientes, relaciones estadísticas y biológicas entre las variables y el tamaño y características de la muestra utilizada en el desarrollo de la ecuación.⁴

a) Un factor fundamental que afectará a la validez global del modelo predictivo es que el método usado para medir la variable que será utilizada como dependiente en la construcción de la ecuación de regresión sea fiable y preciso; si esto no se cumple, la ecuación obtenida tendrá escasa aplicabilidad práctica. En investigación la variable dependiente, la masa libre de grasa en nuestro caso, se mide por técnicas consideradas de referencia; desde hace relativamente pocos años la determinación de la masa libre de grasa por absorciometría de rayos X de doble energía (DXA) se ha ido asentando como método de referencia en muchos estudios,^{14,15} ya que reúne factores como amplia disponibilidad, precisión, fácil entrenamiento del personal que realiza la prueba –lo que reduce las fuentes de error exploratorio– y no producir molestias al paciente. Una limitación que se esgrime contra la DXA es el déficit de concordancia entre los distintos modelos.

Los métodos de referencia que miden la variable dependiente siguiendo modelos multicompartimentales y que además requieran un alto grado de colaboración por parte del paciente son más sensibles a la acumulación de errores en la medición.

b) La precisión en las medidas de las variables independientes o predictoras también es muy importante. Variables como los datos antropométricos y la medida de la impedancia bioeléctrica deben realizarse con una metodología fiable, que minimice el error inter e intraobservador, asegurando en lo posible que la variabilidad se debe realmente a las diferencias existentes entre los individuos. Además, todos los instrumentos empleados deben estar adecuadamente calibrados.

c) Entre las características de la muestra utilizada para el desarrollo de la ecuación debe-

mos citar el tamaño, que dependerá de la correlación entre la variable dependiente y las predictoras; si es elevada, no se requerirá un número elevado de casos, existiendo trabajos en la literatura con ecuaciones publicadas que fueron desarrolladas con muestras que giran en torno a los 100 casos. No obstante, como regla general, cuanto mayor sea el número de casos empleado más fiable será la ecuación desarrollada a partir de la muestra; como se puede comprobar experimentalmente, los incrementos en el valor de R^2 que se producen con la adición de variables predictoras al modelo tienen mayor probabilidad de resultar significativos en muestras que superan los 100 casos.

d) Ya mencionamos las consideraciones estadísticas y los criterios biológicos de las variables que formarán parte del modelo en el apartado anterior, y en cuanto al método de desarrollo de las ecuaciones predictivas también hemos comentado que se trata de una regresión lineal múltiple, que consta de dos modos básicos de seleccionar las variables que formarán parte de la ecuación:¹³

- El modo de *selección hacia delante*, que consiste en ir seleccionando las variables que presenten una mejor correlación con la variable dependiente para incluirlas en el modelo hasta que no se obtengan incrementos significativos en el coeficiente de correlación lineal múltiple R^2 .
- El *modo hacia atrás*, que parte del supuesto inverso al anterior. Se incluyen todas las variables predictoras posibles en el modelo y posteriormente vamos eliminando las variables cuya permanencia no supone mejoría en el modelo, según los criterios especificados en nuestro estudio
- Un tercer modo es el llamado *regresión paso a paso*, que es una modificación del método de selección hacia delante, ya que parte de la inclusión progresiva de variables valorando continuamente la significación del modelo, de tal forma que la inclusión de una variable puede significar la salida de una existente si se reduce la significación de la ecuación que estamos construyendo. El proceso se da por finalizado cuando ninguna variable de las que

está en la ecuación cumple las condiciones para salir, y ninguna de las que no está incluida en la ecuación cumple condiciones para entrar.

Elección del modelo

Una vez escogidas las variables que serán utilizadas para la confección de la ecuación se construirán todas las ecuaciones posibles y, posteriormente, se seleccionará la que cumpla los mejores criterios de ajuste desde el punto de vista estadístico, estos criterios se basan en:⁶

- Mayor valor del *coeficiente de determinación R^2 ajustado*, que representa la proporción de la varianza total de la variable dependiente que es explicada por las variables independientes incluidas en el modelo.
- Menor valor del *error típico de la estimación*.
- Valor del *índice C_p de Mallows*, que representa un indicador del número adecuado de variables predictoras a incluir en el modelo. El cálculo de este índice se basa en el cociente entre la dispersión residual entre cada modelo con p variables y la varianza residual del modelo con todas las variables posibles, teniendo en cuenta el número de casos estudiados, según la siguiente fórmula:

$$C_p = (SCE_p / s^2) - [n - 2(p+1)]$$

siendo SCE_p la suma de cuadrados de los errores con p variables, s^2 la varianza en el modelo con todas las variables, n el número de casos y p el número de variables en cada ecuación.

En general, un modelo adecuado presentará un C_p de Mallows muy próximo al número de variables predictoras incluidas en el modelo. Para ello se puede hacer un gráfico en el cual el eje de ordenadas representa los valores de C_p y el de abscisas el número de variables para cada ecuación posible. El número adecuado de variables predictoras a incluir en el modelo será el situado próximamente a la intersección de las líneas que representan ambos valores, o bien se puede considerar que si C_p está próximo a $p + 1$, el modelo es adecuado en cuanto al número de variables.

Validación de las ecuaciones

Una vez obtenida la ecuación para la variable que deseamos estimar, debemos comprobar su precisión y validez; inicialmente se calcula el coeficiente de variación que es la desviación típica medida en unidades de media. Es un índice adimensional expresado en porcentaje, fundamental en la comparación de métodos de medición clínica y que, al no depender de unidades de medida, nos permite comparar métodos y modelos en diferentes estudios;¹⁶ realizaremos por lo tanto una comparación entre los coeficientes de variación del modelo matemático y de la técnica de referencia en la muestra de validación.

Un método riguroso para validar la ecuación desarrollada y comprobar su fiabilidad es la validación cruzada que consiste en aplicar dicha ecuación a una muestra diferente a la utilizada para su desarrollo; alternativamente, cuando partimos de una muestra de datos lo suficientemente amplia, podemos seleccionar de forma aleatoria una submuestra previamente al desarrollo de la ecuación para, posteriormente, emplearla como muestra para la validación del modelo.^{3,4}

La validación del modelo se puede expresar mediante el llamado índice de reducción de la validación cruzada, que consiste en hallar la diferencia entre los cuadrados de los coeficientes de correlación lineal obtenidos entre los valores estimados por la ecuación y los valores realmente medidos de la variable dependiente en ambas muestras, la empleada para el desarrollo de la ecuación y la muestra para validación; se ha establecido que valores de este índice inferiores a 0,10 significarían que la ecuación obtenida es muy fiable, mientras que valores superiores a 0,90 indicarían una muy baja fiabilidad de la ecuación para ser aplicada en individuos que no pertenezcan a la muestra utilizada para su desarrollo.

El error típico de la estimación obtenido al desarrollar el modelo de regresión se emplea como un índice que indica el grado de ajuste de la ecuación y, además, se emplea como valor para comparar el llamado error puro (EP),^{4,6} que se obtiene de la raíz cuadrada de la suma de las diferencias al cuadrado entre el valor de la variable dependiente estimado por la ecuación y el realmente medido, ambos en la muestra de validación, dividido por el nú-

mero de casos de la muestra de validación, como se muestra en la siguiente fórmula. Lo deseable es que ambos valores sean similares; en caso de que el valor del error puro sea notablemente mayor significa que la ecuación obtenida, en una determinada muestra y con un buen ajuste, no es aplicable en otros casos; las causas tendremos que buscarlas en las variables empleadas, características de la muestra como edades, sexo, etc.

$$EP = \sqrt{\frac{\sum (\tilde{Y} - Y)^2}{n}}$$

siendo \tilde{Y} el valor de la variable dependiente estimado por la ecuación, Y el valor de la variable dependiente realmente observado y n el tamaño de la muestra de validación.

Otro método para complementar la validación del modelo matemático obtenido es aplicar el análisis gráfico desarrollado por Bland y Altman,¹⁷ que consiste en representar mediante un gráfico las diferencias obtenidas entre dos mediciones realizadas entre dos sistemas de medida, frente a su media.

En nuestro caso, como se trata de validar un método, que es la ecuación, el otro sistema de medida tendrá que ser un sistema considerado de referencia, por ejemplo la DXA. El procedimiento consistiría en aplicar la ecuación obtenida a la muestra de validación para obtener el valor de la masa libre de grasa y hallar las diferencias con el valor obtenido con la DXA en cada individuo para, posteriormente, hallar los valores promedio individuales entre el valor obtenido con la ecuación y el obtenido con la DXA.

Los valores obtenidos se representan en un gráfico de dispersión en el que el eje de abscisas contiene los valores promedio y el eje de ordenadas los valores de las diferencias entre ambos métodos. Un ejemplo de la aplicación de este método lo podemos observar en la Figura 1, en la que se muestra el análisis gráfico efectuado para analizar la concordancia entre una ecuación desarrollada para estimar el peso corporal y el peso realmente medido en una muestra de individuos diferente a la utilizada para el desarrollo de la ecuación.

El gráfico resultante nos proporciona información acerca de:

1. Concordancia entre las mediciones: la validez de la ecuación será tanto mayor cuanto más valores igual a cero o próximos a cero resulten en la diferencia entre ambos métodos (la medición obtenida con la ecuación sería concordante con el método de referencia), lo que se observa por los puntos sobre la línea del valor cero del eje de ordenadas, paralelo al eje de abscisas.
2. En el caso de que no exista buena concordancia, el gráfico nos informará de si existen tendencias, por ejemplo si con la ecuación se obtienen sistemáticamente valores más altos o más bajos que con la técnica de referencia; y si estos valores se obtienen característicamente en ciertos grupos de individuos, por ejemplo, entre los de masa libre de grasa más elevada.
3. El método desarrollado por Bland y Altman proporciona, además, los límites de concordancia obtenidos mediante el intervalo de confianza para la diferencia entre las mediciones efectuadas por ambos métodos. Estos valores se comparan con los que se hayan fijado como aceptables al comienzo del análisis de validación por lo que podremos saber si las diferencias obtenidas al aplicar la ecuación son asumibles, o por el contrario la diferencia respecto a la técnica de referencia invalida la aplicación de la ecuación en individuos diferentes a la muestra obtenida para su desarrollo.

Recientemente otros autores han desarrollado un método gráfico para evaluar la concordancia entre dos métodos de medida,¹⁸ complementando el método propuesto por Bland y Altman. En el caso

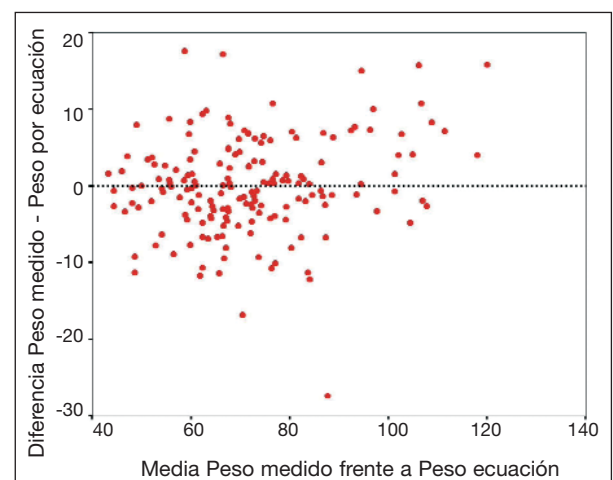


Figura 1. Ejemplo de análisis gráfico.

de la validación de un modelo matemático para estimar la masa libre de grasa, consistiría en representar en el eje de abscisas las diferencias absolutas entre los valores obtenidos aplicando la ecuación y los obtenidos con la técnica de referencia, y en el eje de ordenadas la proporción de valores en los que las diferencias igualan, al menos, cada una de las diferencias observadas; obviamente, se tendrá que emplear una muestra de individuos diferente a la utilizada para el desarrollo de la ecuación.

Las ventajas aportadas por este método se objetivan en que puede aportar más información acerca de la posible relevancia clínica de las diferencias entre los métodos de medida, y que permite realizar análisis de validez y fiabilidad aplicando las técnicas estadísticas de análisis de la supervivencia.

El coeficiente de correlación intraclase, proporciona un valor numérico que expresa el grado de concordancia entre dos valores obtenidos por diferentes métodos de medida aplicados a la misma muestra de individuos. En nuestro caso los métodos de medida serían el método considerado de referencia y la ecuación desarrollada.¹⁹

La base teórica del coeficiente de correlación interclase se basa en descomponer la variabilidad inherente a toda medida en tres componentes, la variabilidad entre individuos y la variabilidad en el mismo individuo representada por los valores que miden los dos métodos empleados, y la variabilidad residual, que representa el componente de error que conlleva todo proceso de medición.

El cálculo del coeficiente de correlación interclase (CCI) se realiza mediante la técnica estadística del análisis de la variancia para medidas repetidas según la fórmula que mencionamos a continuación, y sus valores pueden oscilar entre 0 y 1, considerándose a efectos prácticos que un coeficiente de correlación interclase inferior a 0,30 expresa que no existe concordancia entre los métodos, mientras que, por el contrario, un

coeficiente de correlación interclase superior a 0,90 reflejaría una excelente concordancia; en el caso de la validación de una ecuación de regresión, al ser comparada con un método de referencia, un coeficiente de correlación interclase menor de 0,30 reflejaría que la ecuación no obtiene resultados válidos al ser aplicada a una muestra independiente.

$$CCI = (kSS_B - SS_T) / (k - 1)SS_T$$

Siendo k el número de observaciones por sujeto, SS_B la suma de cuadrados entre los sujetos y SS_T la suma de cuadrados totales.

En los casos en que no se dispone de una muestra de individuos lo suficientemente amplia para realizar la validación de las ecuaciones desarrolladas se emplea el método de predicción de la suma de cuadrados (PRESS en la literatura anglosajona), que se basa en desarrollar la ecuación de regresión excluyendo un caso, estimar el valor de este caso con la ecuación desarrollada y calcular su residuo (la diferencia del valor real y el estimado). Repitiendo estos pasos para todos los casos, posteriormente se calculará el valor de la raíz cuadrada de la suma de los cuadrados de los residuales divididos por el número total de casos; si el resultado es un valor próximo al error típico de la estimación obtenido al desarrollar el modelo será indicativo de que la ecuación desarrollada es fiable.

La ecuación desarrollada y validada se considera adecuada si presenta errores de predicción (equivalente al valor del error típico de estimación obtenido al desarrollar la ecuación) menores de 3 kg y 2,8 kg en hombres y mujeres respectivamente, frente a la técnica de referencia. En la Tabla 2 podemos ver algunas de las ecuaciones publicadas en la literatura para la estimación de masa libre de grasa y sus respectivos errores de estimación.

Tabla 2. Ecuaciones publicadas para la estimación de la masa libre de grasa y sus respectivos errores de estimación

Población de estudio	Ecuación	Error de estimación
Adultos sanos ⁵	$-4,104 + 0,518 T^2/R + 0,231 \text{ peso} + 0,130Xc + 4,229 \text{ sexo}$	1,8
Mayores de 16 años sanos ⁸	$-12,44 + 0,34 T^2/R + 0,1534 T + 0,273 \text{ peso} - 0,127 \text{ edad} + 4,56 \text{ sexo}$	2,6
Varones de 12-94 años ⁹	$-10,68 + 0,65 T^2/R + 0,26 \text{ peso} + 0,02 R$	3,9
Mujeres de 12-94 años ⁶	$-9,53 + 0,69 T^2/R + 0,17 \text{ peso} + 0,02 R$	2,9
Adultos sanos de 35-65 años ²⁰	$-14,94 + 0,28 T^2/R + 0,18 \text{ peso} + 0,23 T + 0,064 (\text{peso sexo}) - 0,077 \text{ edad}$	3,6

T: talla en cm, R: resistencia a 50 KHz en ohmios, Xc: reactancia en ohmios.

Bibliografía

1. Wang ZM, Heshka S, Pierson RN, Jr., Heymsfield SB. Systematic organization of body-composition methodology: an overview with emphasis on component-based methods. *Am J Clin Nutr* 1995; **61**: 457-65.
2. Armitage P, Berry G. Mediciones múltiples. Estadística para la investigación biomédica. Harcourt Brace España: Madrid, 1997; 295-316.
3. Guo SS, Chumlea WC, Cockram DB. Use of statistical methods to estimate body composition. *Am J Clin Nutr* 1996; **64**: 428S-35S.
4. Guo SS, Chumlea WC. Statistical methods for the development and testing of predictive equations. En: Roche AF, Heymsfield SB, Lohman TG, eds. Human body composition. Champaign, IL 61825-5076: *Human Kinetics* 1996: 191-202.
5. Kyle UG, Genton L, Karsegard L, Slosman DO, Pichard C. Single prediction equation for bioelectrical impedance analysis in adults aged 20-94 years. *Nutrition* 2001; **17**: 248-53.
6. Sun SS, Chumlea WC, Heymsfield SB, Lukaski HC, Schoeller D, Friedl K, et al. Development of bioelectrical impedance analysis prediction equations for body composition with the use of a multicomponent model for use in epidemiologic surveys. *Am J Clin Nutr* 2003; **77**: 331-40.
7. Segal KR, Van LM, Fitzgerald PI, Hodgdon JA, Van Itallie TB. Lean body mass estimation by bioelectrical impedance analysis: a four-site cross-validation study. *Am J Clin Nutr* 1988; **47**: 7-14.
8. Deurenberg P, van der KK, Leenen R, Weststrate JA, Seidell JC. Sex and age specific prediction formulas for estimating body composition from bioelectrical impedance: a cross-validation study. *Int J Obes* 1991; **15**: 17-25.
9. Stolarczyk LM, Heyward VH, Hicks VL, Baumgartner RN. Predictive accuracy of bioelectrical impedance in estimating body composition of Native American women. *Am J Clin Nutr* 1994; **59**: 964-70.
10. Kotler DP, Burastero S, Wang J, Pierson RN, Jr. Prediction of body cell mass, fat-free mass, and total body water with bioelectrical impedance analysis: effects of race, sex, and disease. *Am J Clin Nutr* 1996; **64**: 489S-97S.
11. Roubenoff R, Baumgartner RN, Harris TB, Dallal GE, Hannan MT, Economos CD, et al. Application of bioelectrical impedance analysis to elderly populations. *J Gerontol A Biol Sci Med Sci* 1997; **52**: M129-36.
12. Heitmann BL. Evaluation of body fat estimated from body mass index, skinfolds and impedance. A comparative study. *Eur J Clin Nutr* 1990; **44**: 831-7.
13. Carrasco JL, Hernán MA. Análisis de regresión lineal. En: Carrasco JL, Hernán MA, eds. Estadística multivariante en las ciencias de la vida. Editorial Ciencia 3: Madrid, 1993; 153-95.
14. Pietrobelli A, Heymsfield SB, Wang ZM, Gallagher D. Multi-component body composition models: recent advances and future directions. *Eur J Clin Nutr* 2001; **55**: 69-75.
15. Kyle UG, Bosaeus I, De Lorenzo AD, Deurenberg P, Elia M, Gómez JM, et al. Bioelectrical impedance analysis—part I: review of principles and methods. *Clin Nutr* 2004; **23**: 1226-43.
16. Martín Andrés A, Luna J. Estadística descriptiva. En: Martín Andrés A, Luna J, eds. Bioestadística para las ciencias de la salud. Madrid: Ediciones Norma S.A. 1993: 38.
17. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986; **1**: 307-10.
18. Luiz RR, Costa AJ, Kale PL, Werneck GL. Assessment of agreement of a quantitative variable: a new graphical approach. *J Clin Epidemiol* 2003; **56**: 963-7.
19. Prieto L, Lamarca R, Casado A. La evaluación de la fiabilidad en las observaciones clínicas: el coeficiente de correlación intraclase. *Med Clin Barc* 1998; **110**: 142-5.
20. Heitmann BL. Prediction of body water and fat in adult Danes from measurement of electrical impedance. A validation study. *Int J Obes* 1990; **14**: 789-802.